#1947F

Effect of Formalin Fixation on Targeted DNA Sequencing and SNV Identification

Ravi Alla[1], Shujun Luo[1], Elena Helman[1], Michael James Clark[1], Sean Michael Boyle[1], Scott Kirk[1], Parin Sripakdeevong[1], Mirian Karbelashvili[1], Deanna Church[1], Michael Snyder[1, 2], John West[1], Richard Chen[1]

[1]Personalis, Inc., Menlo Park, CA; [2]Department of Genetics, Stanford University, Palo Alto CA.

Contact:
ravi.alla@personalis.com

Introduction

Tumor biopsies are often Formalin-Fixed and Paraffin-Embedded (FFPE) for histological staining, genetic testing and archival purposes. Formalin treatment preserves tissue by crosslinking proteins, but also leads to mutation of the nucleic acid bases and poses a challenge to identication of true variants in the tumor using next-generation sequencing (NGS) methods. Studies have shown that it is possible to isolate high quality nucleic acids from good FFPE samples and to profile small variants using NGS. These studies used tissue fixed with 10% neutral buered formalin for 24 hours, a standard protocol in the pathology field. In our initial handling of FFPE samples we found that the quality of the isolated DNA and subsequent sequencing results vary widely. We hypothesized that this may be due to deviations from the standard protocol, such as inaccurate logging of the fixation protocol, variation in the fixation time, and varied storage conditions of the samples. To understand the role of formalin fixation on the quality of variants called, we performed an augmented target enrichment and sequencing assay on fresh frozen (FF) and formalin treated reference cell line NA12878. We assessed raw DNA quality, library quality, sequencing metrics (alignment rate, duplication rate and on-target eciency) and variant concordance profiles between FF and formalin treated cells. Using the results of this study, we intend to provide guidance to labs preparing FFPE samples about conditions that minimize formalin- induced variations. We also demonstrate how a deeper understanding of the effects of formalin can improve sequencing analysis results from formalin fixed tissues, especially at lower allele frequencies where formalin-related errors have the greatest impact.

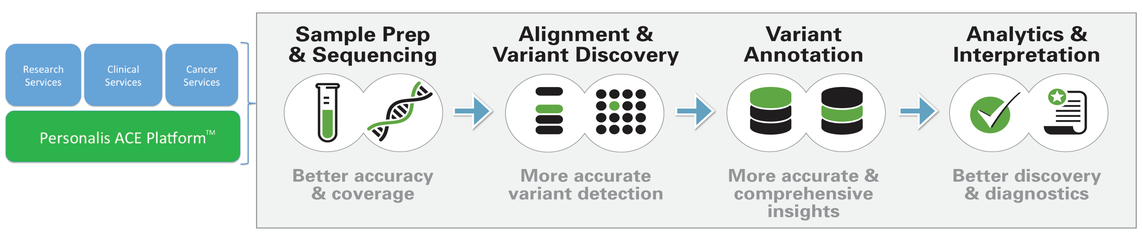
Experiment Outline

NA12878 cell line was subjected to formalin treatment using various protocols, as per the matrix in Table 1. Protocols varied fixation time, fixation temperature and the fixation buffer conditions (which mimic fresh buffered vs old stored formalin solutions). Following sample and library prep, Personalis ACE Cancer Plus Enrichment, sequencing and variant discovery/ annotation (Figure 1) were performed on the orange shaded samples (Table 1)

Table 1: Formalin Fixation Conditions for NA12878 Cell Line

Fixation Time	1 Day			3 Days		
Fixation Temp	RT	37C	45C	RT	37C	45C
Fixation Buffer %PBS	0	0	0	0	0	0
Fixation Buffer %PBS	20	20	20	20	20	20
Fixation Buffer %PBS	40	40	40	40	40	40
Fixation Buffer %PBS	60	60	60	60	60	60
Fixation Buffer %PBS	80	80	80	80	80	80
Fixation Buffer %PBS	100	100	100	100	100	100

Figure 1: Personalis ACE Cancer Plus Pipeline



Formalin Damage Affects DNA and Library Qualities

Quality of DNA isolated from formalin treated NA12878 cells varies considerably between different protocols. Traditional formalin fixation protocol for 1 Day in 100% PBS buffered formalin at RT yields DNA that is high molecular weight and not fragmented (asterisk, Figure 2). The library constructed from this preparation is also of good quality (asterisk, Figure 3). Fixation protocols that use unbuffered formalin at higher temperatures and for prolonged periods of time have highly fragmented low molecular weight DNA (daggers, Figure 2). This poor DNA quality consequently led to poor library efficiency (daggers, Figure 3). Based on these results we believe that fixation using buffered formalin and at the right temperature is crucial for viable DNA for downstream NGS protocols.

Figure 2: Raw DNA Size Distribution of Formalin Treated NA12878 Cells

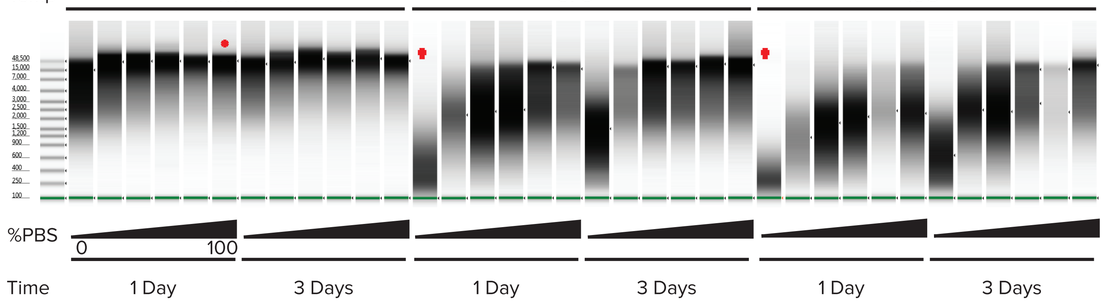
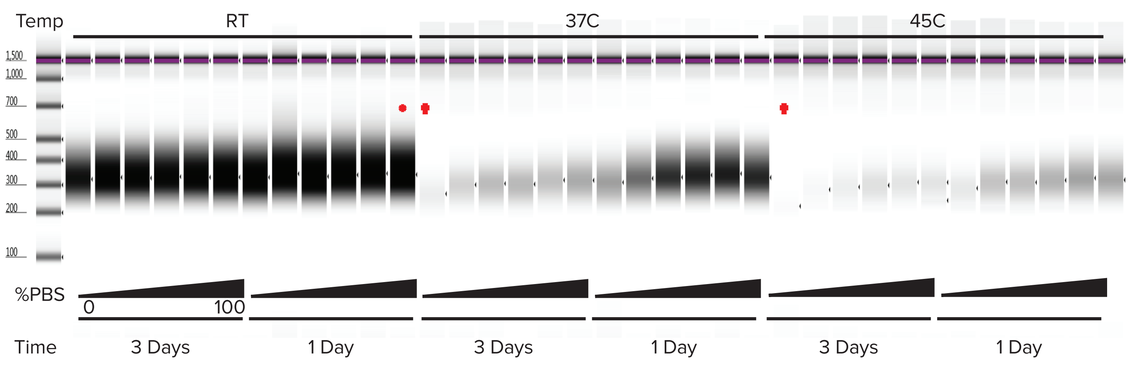


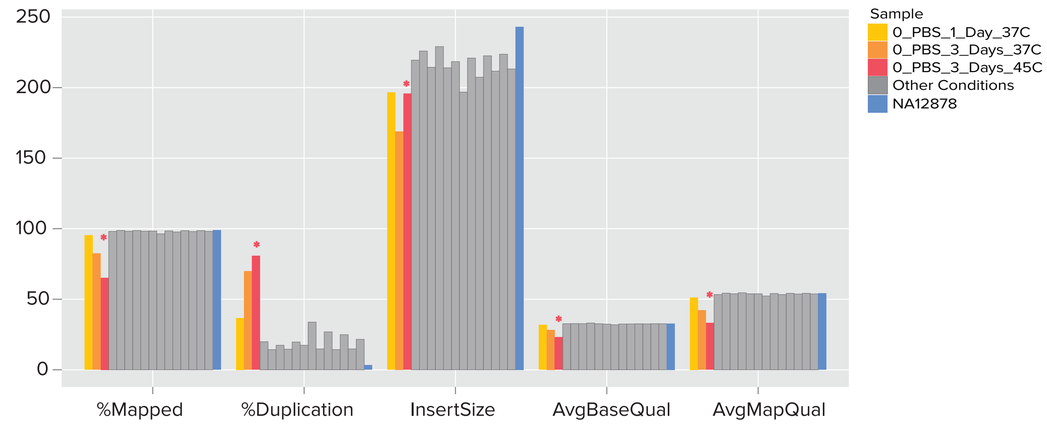
Figure 3: Next Generation Sequencing Library Sizes of Formalin Treated NA12878 Cells



Severely Formalin Damaged Libraries have Low Alignment Metrics

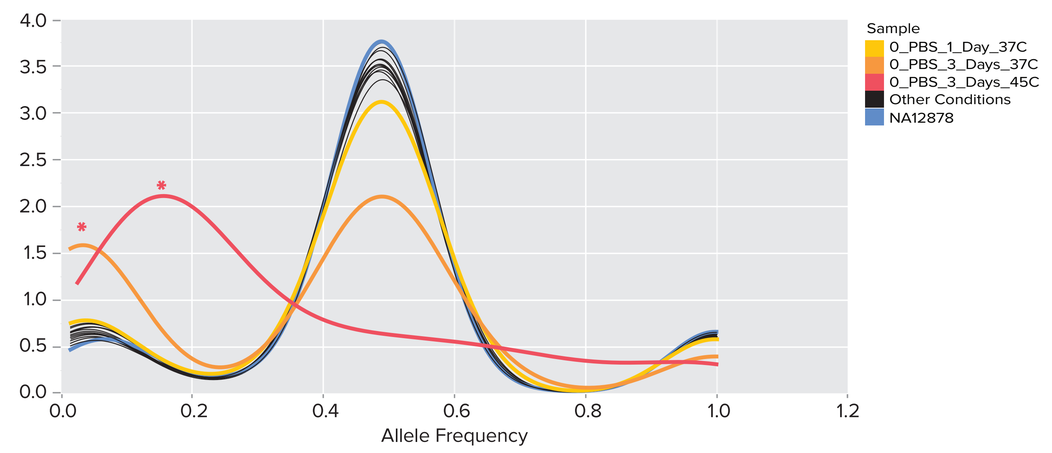
Formalin damaged libraries (samples shaded orange in Table 1) were sequenced on Illumina Hiseq and the resulting data were aligned to human reference assembly hs37d5. Highly damaged formalin samples had poor mapping rates, poor mapping qualities, high duplication rates, lower insert sizes (owing to more fragmented starting DNA) and lower recalibrated (GATK) base quality scores (Figure 4 red asterisk).

Figure 4: Alignment Metrics for FF vs Formalin treated NA12878 Cells



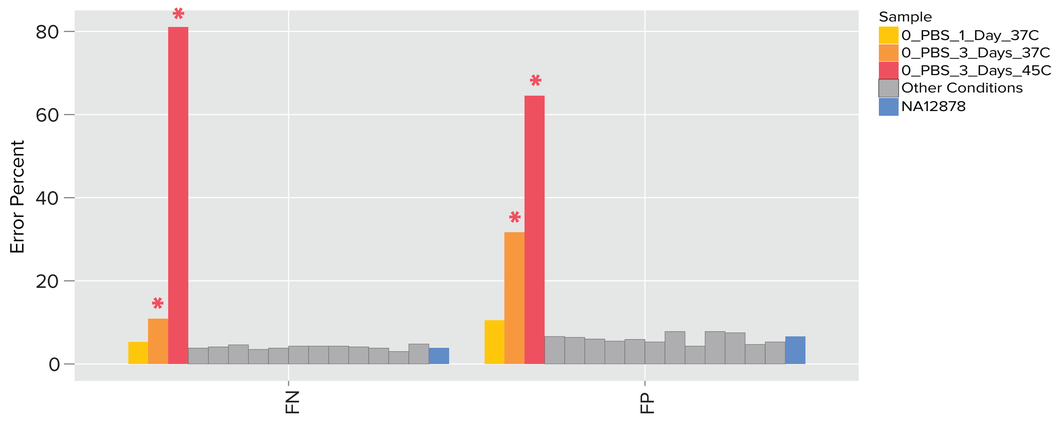
Formalin Damage Increases Low Allele Frequency Somatic Variants

Figure 5: Allele Frequency Distribution of Somatic Variants in FF vs Formalin treated NA12878 Cells



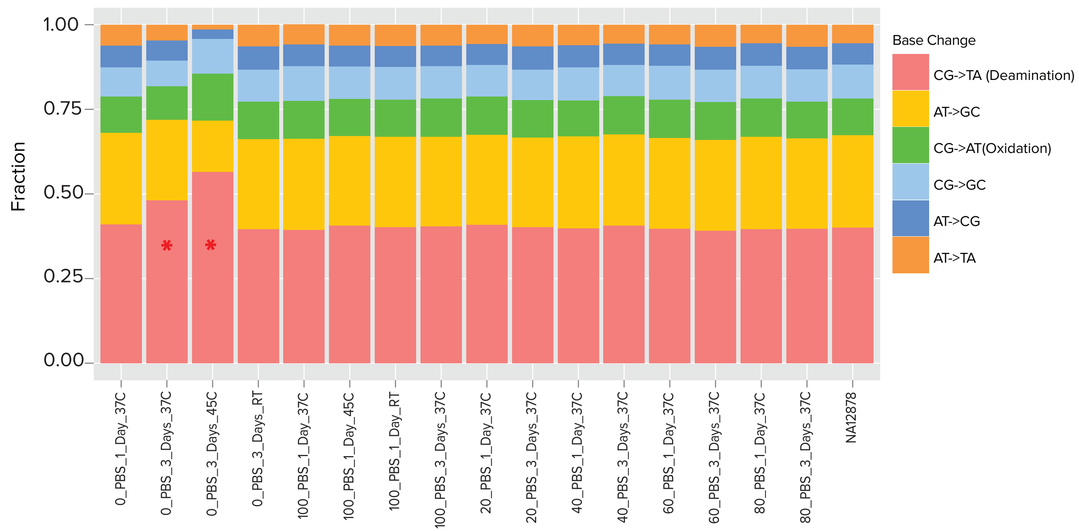
Formalin Damage Decreases Somatic Variant Calling Accuracy

Figure 6: False Positive and False Negative Errors in Formalin Treated vs FF NA12878 Cells



Formalin Causes Increased Deamination (C->T and G->A) Mutations

Figure 7: Somatic Variant Nucleotide Change Context in Formalin Treated vs FF NA12878 Cells



Summary

Protocols that use long periods of fixation at high temperatures in unbuffered Formalin adversely affect tissue DNA. The DNA is more fragmented, has lower library preparation efficiency and leads to reduced sensitivity and specificity in somatic variant calling. Pathology labs which fix human tissue for histology and storage should adhere to proper fixation protocols, so that the tissue may be used in NGS diagnostic assays.

Personalis®
Pioneering Genome-Guided Medicine